RM-69-22

AD695411

AN INVESTIGATION OF THE ACCURACY OF THE PEARSON SELECTION FORMULAS

Melvin R. Novick

and

Dorothy T. Thayer

Office of Naval Research Contract 000-14-69C-0119

Melvin R. Novick, Principal Investigator

Educational Testing Service

Princeton, New Jersey

September 1969

49

RM-69-22

AN INVESTIGATION OF THE ACCURACY OF THE PEARSON SELECTION FORMULAS

Melvin R. Novick
and
Dorothy T. Thayer

Office of Naval Research Contract 000-14-69C-0119
Melvin R. Novick, Principal Investigator

Educational Testing Service
Princeton, New Jersey

September 1969

An Investigation of the Accuracy of the Pearson Selection Formulas

Melvin R. Novick and Dorothy T. Thayer

Introduction

The Pearson formulas for correcting correlation coefficients for restriction of range are based on crucial assumptions of linearity of regression and homoscedasticity of the error distributions. Some small studies, of which that of Rydberg (1963) is the most comprehensive, have previously been undertaken to determine the accuracy of these formulas. The general result found by Rydberg and others previously is that for small or moderate degree of selection the Pearson formulas are reasonably accurate but with some tendency to undercorrect. The present study was designed to investigate the accuracy of these formulas both for moderate and for extreme degrees of selection and to do so on many different types of variables. The unique feature of the present study is the very large sample sizes available for each of the data sets. With sample sizes of approximately 20,000 cases it is possible to perform extreme selection and still maintain relatively large sample sizes in the selected group. Thus investigations in these restricted subpopulations will not suffer from overly erratic fluctuation because of small sample sizes. While it is too much to say that in the case of extreme selection, we can treat the sample correlations within any extreme selection group as the true population correlation, still sampling variation should not appreciably distort our findings.

The central importance of a correction for restriction of range is apparent on noting that when comparing two tests, for example, a new test and an old test, as predictors of some criterion it is seldom possible to obtain criterion correlations for the total applicant group. Almost always selection must continue on

the basis of the old test and a valid comparison between tests cannot be made unless an accurate correction for restriction of range is available. If a correction is not made the general tendency will be to show the old test in a very unfavorable light, and thus to suggest the replacement of the old test with a new test when in fact such action is completely unwarranted.

In the present paper we restrict ourselves entirely to the case in which there exists a well defined explicit selection variable. Our purpose is to pin down as accurately as possible the range in which the Pearson formulas are acceptable both for explicit and incidental selection and to suggest other methods for cases in which they are not. A major problem in the application of range restriction corrections is the difficulty in isolating the actual selection variable. In most applications in which test scores are used for selection, they are not used on an exclusive basis so that, in fact, many other variables enter into selection. A popular way of "handling" this problem is to use a multivariate selection formula bringing in data on many incidental selection variables. The efficacy or even the logical justification for this approach has never been demonstrated. Moreover it should be clear that such a technique can be valid only to the extent that the simpler univariate and bivariate explicit and incidental selection formulas are valid when selection has, in fact, been explicit. Thus we are thrust back to the fundamental task of evaluating the simplest selection formulas. If we are to aspire to a personnel technology, as opposed to a personnel alchemy, we must be sure that popular corrections really do provide the needed corrections.

Description of Data Sets

Two major data sets were used in this study. The first of these was that used by Halpern to obtain norms for the Preliminary Scholastic Aptitude Test (PSAT) and the Academic Interest Measures (AIM). Halpern's data consisted of test scores on approximately 60,000 students in 180 schools. These students had taken the PSAT, the AIM and had completed a student questionnaire. The PSAT provided a verbal aptitude and a mathematical aptitude score. The AIM provided measures of interest in Biological Sciences, English, Fine Arts, Mathematics, Social Sciences, Secretarial, Physical Sciences, Foreign Languages, Music, Engineering, Home Economics, and Executive Occupations. Data from the student questionnaire were not used in the present study. The PSAT-AIM data consisted of scores on approximately 21,000 sophomores, 20,000 juniors and 18,000 seniors. It was decided that for the present study we would limit ourselves to the juniors. For this group data were available on 19,584 students. However data on many students were incomplete. Therefore for convenience it was decided to base our analysis only on those students who had complete scores on all PSAT and AIM scales. Data on 17,001 such students were available. The PSAT-V score is a scale score based on responses to the 70 PSAT-V items. Similarly PSAT-Q score gives a scaled score in the range 20-80 based on the marks on the 50 PSAT-Q items. Specifically, in each case the final score is obtained by taking the number of correct responses, subtracting a percentage of incorrect responses to get a formula score, and then linearly scaling into the interval 20-80. Each of the 12 AIM scale scores is based on responses to 16 keyed items of the AIM inventory. The reported scores are scaled so as to lie in the range 0-32.

Distributions of Test Scores

Table 1 gives the approximate means and standard deviations for the group on each of the two PSAT scales and 12 AIM scales. Table 2 gives the univariate

Insert Tables 1 and 2 about here

distributions of the PSAT-V and Q scores together with the percentage of the population at each score level and the cumulative percentage to that level.

A cursory inspection of Table 2 indicates that both PSAT distributions are positively skewed for this sample. This can be verified by noting that the coefficients of skewness of the distributions are .57 and .67 while a value of 0 would indicate a symmetric distribution. Since the mean of the PSAT-V scores is far below the center of possible values and nearly 7% of the scores are at the lowest attainable score, 20, it is clear that for the PSAT-V scale there is in fact some floor effect. The fact that a score of 80 was not attained on either scale indicates that no ceiling effect was present. On the whole then the two tests were somewhat difficult for the population of examinees. The coefficients of kurtosis were also computed and found to be -.26 and -.27 indicating that each of the distributions was platykurtic.

Table 3 gives the distributions of each of the AIM scales together with the computed coefficients of skewness and kurtosis. It is clear that for the most part we do not have either symmetric or mesokurtic distributions. We find

Insert Table 3 about here

for Home Economics, Secretarial, Foreign Language, and Executive scales a definite ceiling effect; thus,.there is a tendency towards negative skewness and substantial platykurtosis.

A primary interest of this study was to determine the degree of linearity and homoscedasticity to be found typically among psychological variables. In order to conveniently investigate these aspects for the PSAT-Q and PSAT-V bivariate distribution it was decided to group the data into 21 small class intervals on each of the variables. The resulting bivariate plot is given in Table 4.

Insert Table 4 about here

Table 5 gives the means and standard deviations for each variable when the group is restricted to one of the class intervals on the second variable. In

Insert Table 5 about here

making the computations for this particular table, we have worked from the data in Table 4 and taken each person's score to lie at the midpoint of the class interval in which he falls. While the resulting computations for this table will have some degree of inaccuracy this should not be great because of the smallness of the class intervals. The obvious and important findings from this table are that the two regression lines tend to be reasonably linear except in the very extreme ranges but that the scedastic functions are not at all constant. Thus apparently one of the necessary assumptions of the use of the Pearson formulas is reasonably well satisfied except for extreme selection while the second assumption is not.

In order to facilitate processing of the PSAT data it was decided to further group the data into class intervals on each of the V and Q scales so that as nearly as possible each interval on each scale contains 10% of the population. Table 6 gives a bivariate plot of the PSAT-V and Q scores grouped into these class intervals and the cumulative percentages for each class interval.

Insert Table 6 about here

Table 7 gives the means and standard deviations for each variable when restricted subpopulations are defined by class intervals on the other variable.

Insert Table 7 about here

A cursory inspection of this table makes clear the relative acceptability of the linearity assumption and the complete unacceptability of the homoscedastic assumption. We do not give bivariate plots of the AIM scales with the PSAT scales since it is clear that there is a greater degree of nonnormality in the AIM scales. Thus the assumptions of linearity and homoscedasticity are even less likely to be satisfied.

The second data set used in this study was furnished by Dr. Milton H. Maier of the United States Army Behavioral Science Research Laboratory. Data were furnished on approximately 23,000 subjects; however, data on some subjects were incomplete. Only those subjects with complete information were used giving us 22,172 subjects. Due to technical difficulties 40 cases were lost during processing so that the majority of the results are reported for a total of 22,132 cases. Data on each subject consisted of 11 test scores on the Army Classification Battery, 14 test scores on the Army Differential MOS Battery, the MOS number of the training course to which the subject had been assigned and his final grade in that course.

The Army Classification Battery consists of 11 scales. The names of these scales and the number of items on which they are based are given in Table 8.

Insert Table 8 about here

Thirteen of the scales were taken from the Army Differential MOS Battery. The names of these scales and the numbers of items on which each is based are given in Table 9.

Insert Table 9 about here

Table 10 gives the means and standard deviations of each of the ACB and Differential MOS scales in the applicant group. Table 11 gives the bivariate

Insert Tables 10 and 11 about here

plot for ACB-V and ACB-A when the data have been grouped into 15 class intervals on each of the V and A scales. Table 12 gives the mean and standard deviation

Insert Table 12 about here

of the V and A scores when each of these variables has been restricted to one of the class intervals on the other variable. For these data the homoscedasticity assumption seems better satisfied than in the PSAT data. Coefficients of skewness and kurtosis were computed for both the V and A scales. Coefficients of skewness are -.14 and -.15 and the coefficients of kurtosis are -.36 and -.21.

Experimental Method

The first analysis used the PSAT-AIM data. We performed explicit selection on PSAT-V and assumed incidental selection on PSAT-Q and the 12 AIM variables. The strategy employed was to actually select on PSAT-V variable, determine the relevant correlations in the restricted population using the Pearson formula to correct for restriction of range and then to compare these adjusted values with

the actual correlations in the applicant group. Initial computations then involved determining means, standard deviations, and intercorrelations of the variables in the subpopulations defined by outselection on the left on PSAT-V in the groups 20-21, 20-25, 20-28, 20-34, 20-37, 20-41, 20-45, 20-51, and 20-56. In addition these computations were made for the applicant group, PSAT-V score 20-80.

The corrected correlation matrices and standard deviations were calculated assuming: there was explicit selection on variable X, PSAT-V; the variance of X, and the correlations and intercorrelations are known for the selected group and only the variance of X is available for the applicant group. The formulas used are given below. In these formulas small letters refer to the selected group, and capital letters refer to the applicant group.

$$R_{XY} = \frac{r_{xy} S_X}{\sqrt{s_x^2 - s_x^2 r_{xy}^2 + S_X^2 r_{xy}^2}} \tag{1}$$

$$R_{YZ} = \frac{s_x^2 [r_{yz} - r_{xy} r_{xz}] + r_{xy} r_{xz} S_X^2}{\sqrt{[s_x^2 - s_x^2 r_{xy}^2 + r_{xy}^2 S_X^2][s_x^2 - s_x^2 r_{xz}^2 + r_{xz}^2 S_X^2]}} \tag{2}$$

$$S_Y = s_y \sqrt{1 - r_{xy}^2 + r_{xy}^2 \frac{S_X^2}{s_x^2}} \tag{3}$$

where R_{YZ} and r_{yz} are the correlations between two incidental selection variables,

R_{XY}, r_{xy} and r_{xz} are the correlations of the explicit selection variable with an incidental selection variable,

S_X and s_x are the standard deviations of the explicit selection variable, and

S_Y and s_y are the standard deviations of the incidental selection variables.

A similar procedure was used with the Army data. Explicit selection was made on ACB-V called the X variable and incidental selection was assumed for the other seven ACB variables. Again the analysis assumed that the correlations and intercorrelations were known for the selected group and the variances of X were available for both the selected and applicant groups.

Comparison of the corrected correlations from restricted population with the values from the applicant group did not show the Pearson formulas in good light particularly when selection was at all severe. This is documented in the next section. In an attempt to discover a more generally useful correction, particular attention was given to the scedastic functions. This was done because it was found that the failure of this function to be constant was the primary violation of assumptions exhibited by both sets of data, though more so for PSAT data than the Army data. While several techniques were studied only two showed any promise and only these techniques are reported on here. These techniques involved discarding the assumption of constant error variance and using the assumption that the error variances have a general linear form. Attempts were then made to estimate the parameters of this linear relationship and thus to estimate the residual variance in the total population and to use this to obtain an improved correction for restriction of range.

Analysis of the PSAT-AIM Data

To evaluate the accuracy of the Pearson selection formulas with respect to the PSAT-AIM data, explicit selection was performed on PSAT-V with uccessive percentages in the selected group being approximately 10, 20, 30, 40, 50, 70, 80 and 90. The general pattern of results is illustrated in Table 13.

Insert Table 13 about here

The extrapolated correlations between PSAT-V and PSAT-Q are consistent with previous findings. The correlations (.36, .49, .58, .65, .73) in the successive groups with increasing percentages of selection are extrapolated to the values (.66, .68, .71, .73, .75). Since the true total group correlation is .75 this is certainly a clear and meaningful improvement, though, as found in previous studies, there is a tendency to undercorrect. Despite this substantial correction, however, one can question whether the correction is really adequate.

Suppose PSAT-V is the standard predictor of PSAT-Q and suppose that in current practice the selected group is 50% of the applicant group. Suppose further that a new predictor is being proposed and that this new predictor has a very low correlation with PSAT-V. (Actually this last assumption is most unlikely to occur in practice. We would be most fortunate if it did.) Then the restriction effect on the new predictor would be very small. In applications such as this an increase of .05 in the correlation coefficient would be considered a major advance, yet the Pearson extrapolated validity for the "old" test (PSAT-V) is .04 less than the actual total population value. Clearly in such a case there can be little justification for having any faith in the analysis. To compound the problem further one needs only note that for all other variables the typical result has been an overcorrection.

The fact that the correction works substantially less well with the AIM scales is a clear reflection of the sensitivity of the correction formula to the linearity and homoscedasticity assumptions. These scales are based on fewer items than are the PSAT scales, therefore asymptotic normality and hence linearity and homoscedasticity are much less evident.

It is also worth noting that there is a definite tendency for the correction formulas to be more accurate when the correlation in the applicant population

is substantial. When this correlation is near zero, the formulas seem to be of almost no value.

To evaluate the accuracy of the incidental selection formula we studied the correlations and extrapolated correlations between PSAT-Q and nine of the AIM variables. The results of this analysis are given in Table 14.

Insert Table 14 about here

Analysis of the Army Data

To evaluate the accuracy of the Pearson selection formulas with respect to the Army data explicit selection was performed on ACB-V with successive percentages in the selected group being approximately 10, 20, 40, 60, 80 and 90. The general pattern of results obtained can again be illustrated by looking at a few selected results given in Table 15.

Insert Table 15 about here

The extrapolated correlations between ACB-V and ACB-A, in this case, contradict previous findings. The correlations (.16, .27, .38, .47, .55, .58) in the successive groups with increasing percentages of selection are extrapolated to the values (.42, .60, .64, .65, .65, .64). Since the applicant group correlation is .60 this is again a clear and meaningful improvement. But in this case there has been a nontrivial overcorrection of even greater magnitude than the undercorrection in the PSAT-AIM data and again the correction with respect to all other variables has been an overcorrection.

To evaluate the accuracy of the incidental selection formula we examined Table 16 which gives the correlations and extrapolated correlations between ACB-A and six of the ACB scales. Here again the results were generally unsatisfactory.

Insert Table 16 about here

New Methods for Correcting Correlation Coefficients

Since the assumption of homoscedasticity of the error distribution does not appear to be satisfied for either data set, we attempted to find a procedure which would take into account the heteroscedasticity of errors. Using the linearity of regression assumption we have the following

$$R_{XY} = r_{xy} \frac{s_y S_X}{s_x S_Y} \quad . \qquad (4)$$

The standard deviations of the explicit and incidental selection variables, s_x and s_y, in the selected group and the standard deviation of the explicit selection variable, S_X, in the applicant group are known. The standard deviation of the incidental selection variable, S_Y, in the applicant group has to be estimated to correct the correlation coefficient, r_{xy}, for restriction of range. Two methods of estimating S_Y were attempted.

New method 1 used the analysis of variance breakdown of total variance into the sum of (i) average within-class variance and (ii) among-class variance.

$$\sigma^2(Y) = \mathcal{E}\{\sigma^2(Y|x)\} + \sigma^2\{\mathcal{E}(Y|x)\} \quad . \qquad (5)$$

Thus, to estimate the variance of the incidental selection variable in the applicant group we had to estimate the expected value of the conditional

variances and the variance of the conditional means. Specifically this meant estimating $\sigma^2(Y|x)$ and $\mathcal{E}(Y|x)$ for those values of x in the rejected group.

The applicant group was divided into a selected group and a rejected group by selecting on an explicit selection variable. In the selected group, the conditional means and variances for the incidental selection variable were known for a number of intervals. These conditional means and variances were assumed to have a general linear form over these ordered intervals. By using least squares a straight line was fitted to the known conditional means of the incidental selection variable. The least squares estimates of the slope and intercept were used to obtain by extrapolation an estimate of the conditional mean for the incidental selection variable in the rejected group. The same procedure was used to estimate the conditional variance of the incidental selection variable in the rejected group. A weighted average of the known and estimated conditional variances of the incidental selection variable was used as a pooled estimate of the average conditional variance. The weights used were the number of persons in each class interval. Thus, an estimate of the first term in (5) was obtained.

The second term in (5) was estimated by using the relationship

$$\sigma^2\{\mathcal{E}(Y|x)\} = \mathcal{E}\{\mathcal{E}(Y|x)\}^2 - \{\mathcal{E}(Y)\}^2 \quad . \tag{6}$$

An estimate for the average value of the incidental selection variable was calculated by pooling the known and estimated conditional means of the incidental selection variable. A similar procedure was used to estimate the average value of the squared conditional mean for the incidental selection variable. Hence, by using (6) an estimate of the variance of the conditional means was obtained.

The estimates of the average value of the conditional variance and the variance of the conditional means were combined using (5) to obtain an estimate for the total variance.

This procedure for estimating the total variance of the incidental selection variable was used for each selected group except the first group. By using (4) the corrected correlation coefficients for each selected group were obtained. Tables 17 and 18 give these corrected correlation coefficients.

Insert Tables 17 and 18 about here

New method 2 assumed that the variances of the incidental selection variable in the selected groups had a general linear form. A straight line was fitted to the variances of the incidental selection variable in the selected groups by using least squares. The least squares estimates of the slope and intercept were used to estimate the variance of the incidental selection variable in the applicant group. This procedure for estimating the variance of the incidental selection variable in the applicant group was used for each selected group except the first since at least two points are needed to fit a straight line. The corrected correlation coefficients for each selected group were obtained by using (4); the values are given in Tables 17 and 18.

Summary

The results of this study strongly suggest that corrections for restriction of range are unsatisfactory even for moderate degrees of selection. Initial attempts to develop more sensitive techniques by relaxing the homoscedasticity assumption were not successful but further developments along these lines are

possible. A more promising approach would involve transforming variables, particularly the criterion variable, so as to achieve the required linearity and homoscedasticity.

Reference

Rydberg, S. (1963). Bias in Prediction. Stockholm: Almqvist and Wiksell.

Table 1

Means and Standard Deviations for PSAT and AIM Scales

Scale	Mean	SD
PSAT		
Verbal	36.0	11.2
Quantitative	38.1	11.3
AIM		
English	16.0	8.5
Music	15.0	8.9
Social Sciences	17.0	9.1
Mathematics	14.8	9.8
Physical Sciences	15.6	9.9
Engineering	16.5	9.8
Home Economics	19.2	9.3
Fine Arts	17.3	8.4
Biological Sciences	16.4	8.5
Secretarial	15.4	8.6
Foreign Languages	17.9	10.3
Executive	18.4	7.7

Table 2

Univariate Distributions for PSAT Scales

	PSAT-V			PSAT-Q		
Score	Freq.	Pct.	C-Pct.	Freq.	Pct.	C-Pct.
20	1158	6.8	6.8	134	0.8	0.8
21	362	2.1	8.9	142	0.8	1.6
22	474	2.8	11.7	158	0.9	2.6
23	412	2.4	14.2	343	2.0	4.6
24	448	2.6	16.8	478	2.8	7.4
25	325	1.9	18.7	389	2.3	9.7
26	641	3.8	22.5	797	4.7	14.4
27	808	4.8	27.2	830	4.9	19.2
28	569	3.3	30.6	808	4.8	24.0
29	379	2.2	32.8	734	4.3	28.3
30	420	2.5	35.3	522	3.1	31.4
31	527	3.1	38.4	602	3.5	34.9
32	815	4.8	43.2	702	4.1	39.1
33	537	3.2	46.3	643	3.8	42.8
34	553	3.3	49.6	625	3.7	46.5
35	561	3.3	52.9	424	2.5	49.0
36	523	3.1	55.9	435	2.6	51.6
37	752	4.4	60.4	568	3.3	54.9
38	251	1.5	61.8	482	2.8	57.7
39	452	2.7	64.5	504	3.0	60.7
40	432	2.5	67.0	376	2.2	62.9
41	431	2.5	69.6	386	2.3	65.2
42	476	2.8	72.4	440	2.6	67.8
43	530	3.1	75.5	389	2.3	70.1
44	360	2.1	77.6	387	2.3	72.3
45	316	1.9	79.5	298	1.8	74.1
46	317	1.9	81.3	301	1.8	75.9
47	293	1.7	83.1	385	2.3	78.1
48	276	1.6	84.7	372	2.2	80.3
49	310	1.8	86.5	334	2.0	82.3
50	247	1.5	88.0	270	1.6	83.9
51	245	1.4	89.4	245	1.4	85.3
52	310	1.8	91.2	304	1.8	87.1
53	175	1.0	92.3	266	1.6	88.7
54	159	0.9	93.2	221	1.3	90.0
55	128	0.8	93.9	194	1.1	91.1
56	114	0.7	94.6	161	0.9	92.0
57	163	1.0	95.6	259	1.5	93.6
58	126	0.7	96.3	169	1.0	94.6
59	112	0.7	97.0	148	0.9	95.4

Table 2 (continued)

Score	PSAT-V Freq.	PSAT-V Pct.	PSAT-V C-Pct.	PSAT-Q Freq.	PSAT-Q Pct.	PSAT-Q C-Pct.
60	67	0.4	97.4	73	0.4	95.9
61	86	0.5	97.9	136	0.8	96.7
62	90	0.5	98.4	114	0.7	97.3
63	43	0.3	98.7	81	0.5	97.8
64	42	0.2	98.9	71	0.4	98.2
65	31	0.2	99.1	35	0.2	98.4
66	33	0.2	99.3	40	0.2	98.7
67	42	0.2	99.5	26	0.2	98.8
68	19	0.1	99.6	62	0.4	99.2
69	14	0.1	99.7	54	0.3	99.5
70	11	0.1	99.8	26	0.2	99.7
71	15	0.1	99.9	3	0.0	99.7
72	6	0.0	99.9	3	0.0	99.7
73	4	0.0	99.9	30	0.2	99.9
74	1	0.0	99.9	10	0.1	99.9
75	2	0.0	100.0	5	0.0	100.0
76	7	0.0	100.0	7	0.0	100.0
77	0	0.0	100.0	0	0.0	100.0
78	0	0.0	100.0	0	0.0	100.0
79	1	0.0	100.0	0	0.0	100.0
80	0	0.0	100.0	0	0.0	100.0

	PSAT-V	PSAT-Q
Mean	36.05	38.09
SD	11.20	11.28
Coef. Skewness	.57	.67
Coef. Kurtosis	-.26	-.27

Table 3

Univariate Distributions for AIM Scales

	English			Music		
Score	Freq.	Pct.	C-Pct.	Freq.	Pct.	C-Pct.
0	280	1.6	1.6	510	3.0	3.0
1	238	1.4	3.0	339	2.0	5.0
2	371	2.2	5.2	541	3.2	8.2
3	343	2.0	7.2	414	2.4	10.6
4	469	2.8	10.0	592	3.5	14.1
5	462	2.7	12.7	506	3.0	17.1
6	499	2.9	15.7	635	3.7	20.8
7	524	3.1	18.7	571	3.4	24.2
8	577	3.4	22.1	578	3.4	27.6
9	590	3.5	25.6	602	3.5	31.1
10	648	3.8	29.4	685	4.0	35.1
11	624	3.7	33.1	648	3.8	38.9
12	664	3.9	37.0	708	4.2	43.1
13	667	3.9	40.9	653	3.8	47.0
14	666	3.9	44.8	642	3.8	50.7
15	692	4.1	48.9	596	3.5	54.2
16	698	4.1	53.0	657	3.9	58.1
17	650	3.8	56.8	572	3.4	61.5
18	651	3.8	60.7	610	3.6	65.0
19	611	3.6	64.3	517	3.0	68.1
20	642	3.8	68.0	565	3.3	71.4
21	568	3.3	71.4	485	2.9	74.3
22	583	3.4	74.8	466	2.7	77.0
23	507	3.0	77.8	447	2.6	79.6
24	493	2.9	80.7	450	2.6	82.3
25	463	2.7	83.4	370	2.2	84.5
26	509	3.0	86.4	391	2.3	86.8
27	445	2.6	89.0	338	2.0	88.7
28	428	2.5	91.5	381	2.2	91.0
29	352	2.1	93.6	300	1.8	92.8
30	372	2.2	95.8	380	2.2	95.0
31	273	1.6	97.4	312	1.8	96.8
32	442	2.6	100.0	540	3.2	100.0

	English	Music
Mean	15.99	14.97
SD	8.46	8.89
Coef. Skewness	.06	.20
Coef. Kurtosis	-.94	-.96

Table 3 (Continued)

	Social Science			Math		
Score	Freq.	Pct.	C-Pct.	Freq.	Pct.	C-Pct.
0	344	2.0	2.0	673	4.0	4.0
1	287	1.7	3.7	470	2.8	6.7
2	343	2.0	5.7	756	4.4	11.2
3	383	2.3	8.0	556	3.3	14.4
4	436	2.6	10.5	750	4.4	18.9
5	454	2.7	13.2	597	3.5	22.4
6	447	2.6	15.8	661	3.9	26.3
7	450	2.6	18.5	583	3.4	29.7
8	470	2.8	21.3	661	3.9	33.6
9	487	2.9	24.1	521	3.1	36.6
10	535	3.1	27.3	569	3.3	40.0
11	534	3.1	30.4	463	2.7	42.7
12	568	3.3	33.8	540	3.2	45.9
13	516	3.0	36.8	501	2.9	48.8
14	643	3.8	40.6	541	3.2	52.0
15	700	4.1	44.7	477	2.8	54.8
16	680	4.0	48.7	512	3.0	57.8
17	647	3.8	52.5	508	3.0	60.8
18	667	3.9	56.4	472	2.8	63.6
19	565	3.3	59.7	477	2.8	66.4
20	581	3.4	63.2	396	2.3	68.7
21	543	3.2	66.3	435	2.6	71.3
22	532	3.1	69.5	392	2.3	73.6
23	484	2.8	72.3	404	2.4	76.0
24	473	2.8	75.1	413	2.4	78.4
25	419	2.5	77.6	382	2.2	80.6
26	489	2.9	80.4	381	2.2	82.9
27	482	2.8	83.3	390	2.3	85.2
28	504	3.0	86.2	443	2.6	87.8
29	466	2.7	89.0	414	2.4	90.2
30	590	3.5	92.5	519	3.1	93.3
31	556	3.3	95.7	450	2.6	95.9
32	726	4.3	100.0	694	4.1	100.0

	Social Science	Math
Mean	16.95	14.80
SD	9.06	9.76
Coef. Skewness	-.04	.22
Coef. Kurtosis	-1.04	-1.18

Table 3 (Continued)

Score	Home Economics Freq.	Pct.	C-Pct.	Fine Arts Freq.	Pct.	C-Pct.
0	276	1.6	1.6	173	1.0	1.0
1	228	1.3	3.0	197	1.2	2.2
2	332	2.0	4.9	261	1.5	3.7
3	281	1.7	6.6	310	1.8	5.5
4	369	2.2	8.7	359	2.1	7.6
5	300	1.8	10.5	349	2.1	9.7
6	392	2.3	12.8	425	2.5	12.2
7	341	2.0	14.8	451	2.7	14.9
8	344	2.0	16.8	488	2.9	17.7
9	374	2.2	19.0	470	2.8	20.5
10	383	2.3	21.3	629	3.7	24.2
11	371	2.2	23.5	577	3.4	27.6
12	441	2.6	26.1	633	3.7	31.3
13	458	2.7	28.8	620	3.6	35.0
14	505	3.0	31.7	611	3.6	38.5
15	561	3.3	35.0	627	3.7	42.2
16	614	3.6	38.6	683	4.0	46.3
17	596	3.5	42.2	632	3.7	50.0
18	587	3.5	45.6	704	4.1	54.1
19	515	3.0	48.6	611	3.6	57.7
20	519	3.1	51.7	699	4.1	61.8
21	465	2.7	54.4	642	3.8	65.6
22	490	2.9	57.3	644	3.8	69.4
23	495	2.9	60.2	577	3.4	72.8
24	562	3.3	63.5	592	3.5	76.3
25	542	3.2	66.7	549	3.2	79.5
26	617	3.6	70.3	569	3.3	82.8
27	588	3.5	73.8	523	3.1	85.9
28	801	4.7	78.5	473	2.8	88.7
29	646	3.8	82.3	453	2.7	91.4
30	991	5.8	88.1	498	2.9	94.3
31	872	5.1	93.3	426	2.5	96.8
32	1145	6.7	100.0	546	3.2	100.0

	Home Economics	Fine Arts
Mean	19.20	17.33
SD	9.34	8.45
Coef. Skewness	-.34	-.08
Coef. Kurtosis	-1.03	-.95

Table 3 (Continued)

Score	Physical Science Freq.	Pct.	C-Pct.	Engineering Freq.	Pct.	C-Pct.
0	899	5.3	5.3	589	3.5	3.5
1	554	3.3	8.5	424	2.5	6.0
2	593	3.5	12.0	568	3.3	9.3
3	466	2.7	14.8	469	2.8	12.1
4	531	3.1	17.9	561	3.3	15.4
5	445	2.6	20.5	474	2.8	18.1
6	527	3.1	23.6	506	3.0	21.1
7	456	2.7	26.3	483	2.8	24.0
8	506	3.0	29.3	502	3.0	26.9
9	479	2.8	32.1	446	2.6	29.5
10	501	2.9	35.0	465	2.7	32.3
11	498	2.9	38.0	464	2.7	35.0
12	509	3.0	41.0	452	2.7	37.7
13	462	2.7	43.7	460	2.7	40.4
14	521	3.1	46.7	500	2.9	43.3
15	507	3.0	49.7	463	2.7	46.0
16	660	3.9	53.6	535	3.1	49.2
17	548	3.2	56.8	510	3.0	52.2
18	534	3.1	60.0	587	3.5	55.6
19	488	2.9	62.8	494	2.9	58.5
20	507	3.0	65.8	505	3.0	61.5
21	461	2.7	68.5	454	2.7	64.2
22	447	2.6	71.2	507	3.0	67.2
23	419	2.5	73.6	460	2.7	69.9
24	468	2.8	76.4	499	2.9	72.8
25	414	2.4	78.8	476	2.8	75.6
26	462	2.7	81.5	490	2.9	78.5
27	383	2.3	83.8	504	3.0	81.4
28	449	2.6	86.4	574	3.4	84.8
29	421	2.5	88.9	492	2.9	87.7
30	509	3.0	91.9	625	3.7	91.4
31	501	2.9	94.8	575	3.4	94.8
32	876	5.2	100.0	888	5.2	100.0

	Physical Science	Engineering
Mean	15.61	16.54
SD	9.88	9.85
Coef. Skewness	.06	-.04
Coef. Kurtosis	-1.19	-1.24

Table 3 (Continued)

Score	Biological Sciences Freq.	Pct.	C-Pct.	Secretarial Freq.	Pct.	C-Pct.
0	236	1.4	1.4	80	0.5	0.5
1	242	1.4	2.8	108	0.6	1.1
2	338	2.0	4.8	204	1.2	2.3
3	354	2.1	6.9	219	1.3	3.6
4	449	2.6	9.5	234	1.4	5.0
5	451	2.7	12.2	280	1.6	6.6
6	495	2.9	15.1	338	2.0	8.6
7	476	2.8	17.9	348	2.0	10.7
8	537	3.2	21.0	413	2.4	13.1
9	544	3.2	24.2	431	2.5	15.6
10	581	3.4	27.7	469	2.8	18.4
11	621	3.7	31.3	435	2.6	20.9
12	645	3.8	35.1	490	2.9	23.8
13	612	3.6	38.7	499	2.9	26.8
14	684	4.0	42.7	572	3.4	30.1
15	703	4.1	46.9	609	3.6	33.7
16	747	4.4	51.3	640	3.8	37.5
17	673	4.0	55.2	646	3.8	41.3
18	642	3.8	59.0	646	3.8	45.1
19	595	3.5	62.5	622	3.7	48.7
20	641	3.8	66.3	673	4.0	52.7
21	572	3.4	69.6	580	3.4	56.1
22	573	3.4	73.0	625	3.7	59.8
23	509	3.0	76.0	546	3.2	63.0
24	524	3.1	79.1	573	3.4	66.3
25	486	2.9	81.9	562	3.3	69.7
26	515	3.0	85.0	591	3.5	73.1
27	441	2.6	87.6	610	3.6	76.7
28	475	2.8	90.4	652	3.8	80.6
29	410	2.4	92.8	690	4.1	84.6
30	409	2.4	95.2	896	5.3	89.9
31	347	2.0	97.2	820	4.8	94.7
32	474	2.8	100.0	900	5.3	100.0

	Biological Sciences	Secretarial
Mean	16.40	19.40
SD	8.53	8.56
Coef. Skewness	1.90	-.27
Coef. Kurtosis	-.96	-.96

Table 3 (Continued)

Score	Foreign Languages Freq.	Pct.	C-Pct.	Executive Freq.	Pct.	C-Pct.
0	982	5.8	5.8	102	0.6	0.6
1	439	2.6	8.4	68	0.4	1.0
2	504	3.0	11.3	168	1.0	2.0
3	369	2.2	13.5	139	0.8	2.8
4	382	2.2	15.7	230	1.4	4.2
5	309	1.8	17.6	232	1.4	5.5
6	315	1.9	19.4	336	2.0	7.5
7	314	1.8	21.3	351	2.1	9.6
8	354	2.1	23.3	436	2.6	12.1
9	331	1.9	25.3	486	2.9	15.0
10	371	2.2	27.5	512	3.0	18.0
11	368	2.2	29.6	512	3.0	21.0
12	418	2.5	32.1	575	3.4	24.4
13	364	2.1	34.2	572	3.4	27.8
14	491	2.9	37.1	679	4.0	31.8
15	492	2.9	40.0	630	3.7	35.5
16	619	3.6	43.7	743	4.4	39.8
17	521	2.1	46.7	759	4.5	44.3
18	509	2.0	49.7	777	4.6	48.9
19	467	2.7	52.5	781	4.6	53.5
20	439	2.6	55.0	812	4.8	58.2
21	456	2.7	57.7	683	4.0	62.2
22	518	3.0	60.8	719	4.2	66.5
23	459	2.7	63.5	702	4.1	70.6
24	513	3.0	66.5	704	4.1	74.7
25	467	2.7	69.2	630	3.7	78.5
26	558	3.3	72.5	681	4.0	82.5
27	503	3.0	75.5	571	3.4	85.8
28	611	3.6	79.1	611	3.6	89.4
29	612	3.6	82.7	496	2.9	92.3
30	759	4.5	87.1	518	3.0	95.4
31	578	3.4	90.5	401	2.4	97.7
32	1609	9.5	100.0	385	2.3	100.0

Foreign Languages		Executive	
Mean	17.85	Mean	18.41
SD	10.32	SD	7.74
Coef. Skewness	-.25	Coef. Skewness	-.21
Coef. Kurtosis	-1.19	Coef. Kurtosis	-.78

Table 4

PSAT-V and PSAT-Q Bivariate Distribution

	PSAT-V																					
PSAT-Q	20 21	22 24	25 27	28 30	31 33	34 36	37 39	40 42	43 45	46 48	49 51	52 54	55 57	58 60	61 63	64 66	67 69	70 72	73 75	76 78	79 80	TOTAL
79-80																						0
76-78												2	1	2			1	1				7
73-75								1	1		4	1	4	7	6	6	8	4		3		45
70-72									2	2	2	3	5	4	10	2		2				32
67-69						1	1	1	7	13	5	14	24	21	17	17	13	4	2	1	1	142
64-66						2	1	6	5	13	16	25	19	17	15	14	8	2	2	1		146
61-63				3	4	8	12	19	31	31	40	52	39	30	27	16	11	7		1		331
58-60			1	2	4	12	21	21	37	37	68	68	40	30	25	11	10	3				390
55-57			4	10	8	26	39	55	69	82	81	76	49	51	34	17	9	3	1			614
52-54		2	3	11	32	39	71	97	111	97	100	76	56	48	30	8	5	3	1	1		791
49-51	1		12	13	52	68	72	103	120	102	98	91	42	41	19	9	5		1			849
46-48	1	5	16	30	71	96	127	145	158	127	99	93	40	25	18	2	3	2				1058
43-45	3	8	37	44	103	136	131	165	156	95	87	41	39	17	8	2	1	1				1074
40-42	9	13	51	78	142	166	186	176	138	96	72	42	21	5	7							1202
37-39	26	62	120	135	242	228	212	198	133	77	70	27	13	6	2	2	1					1554
34-36	60	88	165	161	251	211	192	128	105	63	33	20	5	1	1							1484
31-33	194	216	303	257	329	237	189	108	62	30	11	6	5									1947
28-30	345	277	402	249	306	220	106	78	52	12	11	5	3									2064
25-27	448	396	414	251	242	136	71	28	16	8	4	2										2016
22-24	332	210	184	98	81	44	17	8	3	1	1											979
20-21	101	57	62	26	12	7	7	4														276
TOTAL	1520	1334	1774	1368	1879	1637	1455	1339	1206	886	802	644	405	305	219	106	75	32	7	7	1	17001

Table 5

Approximate Means and Standard Deviations for PSAT-V and PSAT-Q

PSAT-V Interval	PSAT-Q Mean	PSAT-Q SD	PSAT-Q Interval	PSAT-V Mean	PSAT-V SD
20-21	27.1	4.3	20-21	24.5	4.9
22-24	28.4	4.9	22-24	25.0	5.2
25-27	30.1	6.0	25-27	26.7	5.9
28-30	32.3	7.0	28-30	28.8	6.9
31-33	34.4	7.4	31-33	30.8	7.1
34-36	36.9	8.2	34-36	34.3	7.7
37-39	39.4	8.5	37-39	36.5	7.9
40-42	42.0	8.5	40-42	39.1	7.7
43-45	44.4	8.8	43-45	41.1	8.2
46-48	46.9	8.8	46-48	43.4	8.2
49-51	48.7	8.7	49-51	45.3	8.7
52-54	51.4	8.6	52-54	46.8	8.6
55-57	53.2	9.1	55-57	49.1	8.8
58-60	55.6	8.0	58-60	50.9	8.3
61-63	57.0	8.3	61-63	52.3	9.0
64-66	60.2	7.7	64-66	55.7	8.1
67-69	61.2	8.2	67-69	58.0	8.1
70-72	62.3	8.4	70-72	57.8	6.7
73-75	60.7	7.0	73-75	62.1	8.1
76-78	67.1	7.3	76-78	59.9	6.6
79-80	68.0	---	79-80	----	---

Table 6

PSAT-V and PSAT-Q Bivariate Distribution for Grouped Data

PSAT-Q	PSAT-V 20-21	22-25	26-28	29-34	35-37	38-41	42-45	46-51	52-56	57-80	TOTAL	C-PCT
59-80				18	21	39	74	185	203	384	924	5.4
55-58		1	8	22	46	65	120	209	143	169	783	10.0
49-54	1	6	17	130	139	197	317	397	233	203	1640	19.6
44-48	3	15	60	218	238	254	359	349	147	100	1743	29.9
40-43	10	26	85	313	276	254	280	227	85	35	1591	39.3
36-39	48	119	206	519	335	280	244	179	40	19	1989	51.0
33-35	85	158	260	486	264	199	138	79	19	4	1692	60.9
28-32	492	539	694	898	347	212	122	49	14	1	3368	80.7
26-27	372	389	332	354	111	41	15	11	2		1627	90.3
20-25	509	406	356	273	59	25	13	3			1644	100.0
TOTAL	1520	1659	2018	3231	1836	1566	1682	1688	886	915	17001	
C-PCt	100.0	91.1	81.3	69.4	50.4	39.6	30.4	20.5	10.6	5.4		

Table 7

PSAT-V and PSAT-Q Means and Standard Deviations for Grouped Data

PSAT-V Interval	PSAT-Q Mean	PSAT-Q SD	PSAT-Q Interval	PSAT-V Mean	PSAT-V SD
20-21	27.2	4.1	20-25	25.6	5.1
22-25	28.7	5.0	26-27	26.9	5.7
26-28	30.6	6.1	28-32	29.4	6.8
29-34	34.2	7.5	33-35	33.3	7.5
35-37	38.0	8.2	36-39	36.0	8.3
38-41	40.7	8.6	40-43	39.7	8.6
42-45	44.1	8.6	44-48	42.9	9.6
46-51	47.8	8.8	49-54	47.0	10.5
52-56	51.8	8.8	55-58	50.8	10.9
57-80	56.9	8.6	59-80	56.6	11.0

Table 8

Names and Number of Items for Each Scale in Army Classification Battery

Scale Name	Number of Items
Verbal Test	50
Arithmetic Reasoning	40
Pattern Analysis	50
Mechanical Aptitude	45
Army Clerical Speed	110
Army Radio Code	150
Shop Mechanics	40
Automotive Information	40
Electronics Information	40
Classification Inventory	125
General Information Test	50

Table 9

Names and Number of Items for Each Scale in Army Differential MOS Battery

Scale Name	Number of Items
Subtraction and Division Test	100
Tool Knowledge Test	20
Electronics Interest	20
Mechanical Interest	20
Clerical Interest	20
General Adjustment	20
Electronics Knowledge Test	20
Mechanical Principles	20
Mathematical Knowledge Test	20
Science Knowledge Test	20
Pattern Analysis Test	20
Bio-Chem Information Test	30
Electronics Pictures Test	20

Table 10

Means, Standard Deviations and Actual Score Range for ACB

and Differential MOS Scales

Scale	Mean	SD	Actual Score Range
ACB			
Verbal Test	108.9	18.6	50-152
Arithmetic Reasoning	106.1	18.8	50-160
Pattern Analysis	105.6	20.3	57-155
Mechanical Aptitude	107.1	17.0	40-160
Army Clerical Speed	109.7	17.2	50-150
Army Radio Code	99.3	27.0	50-150
Shop Mechanics	106.5	17.2	39-154
Automotive Information	106.9	18.7	55-150
Electronics Information	106.4	18.8	40-160
Classification Inventory	101.2	19.1	40-160
General Information	103.0	16.9	59-160
Differential MOS			
Subtraction-Division	37.0	14.8	0-100
Tool Knowledge	12.5	4.0	0-20
Electronics Interest	7.7	3.9	0-20
Mechanical Interest	11.7	3.8	0-20
Clerical Interest	11.2	3.1	0-20
General Adjustment	13.0	3.1	0-20
Electronics Knowledge	9.8	3.7	0-20
Mechanical Principles	9.6	3.6	0-20
Mathematical Knowledge	9.1	4.1	0-20
Science Knowledge	13.0	5.4	0-30
Pattern Analysis	11.6	4.0	0-20
Bio-Chem Information	17.6	5.9	0-30
Electronics Pictures	9.9	4.2	0-20

Table 11

ACB-V and ACB-A Bivariate Distribution for Grouped Data

ACB-A	ACB-V 39-83	84-90	91-97	98-102	103-107	108-110	111-113	114-116	117-118	119-122	123-125	126-129	130-132	133-136	137-160	TOTAL	C-PCT
135-160	1	2	7	10	27	15	18	36	32	54	90	82	147	295	299	1115	5.0
130-134	2	11	17	29	52	22	23	70	85	70	152	125	237	210	166	1271	10.7
126-129	6	21	29	27	72	45	26	97	97	94	122	161	166	174	118	1255	16.4
121-125	7	16	48	35	81	49	75	98	110	111	156	95	144	147	99	1271	22.1
118-120	22	36	67	97	111	78	72	151	110	126	115	131	114	149	58	1437	28.6
115-117	22	43	88	83	129	87	46	163	96	104	86	104	97	80	38	1266	34.3
113-114	33	66	99	77	127	98	27	116	70	75	37	67	55	51	23	1021	38.9
109-112	49	55	110	88	149	89	83	110	72	110	85	47	65	51	33	1196	44.3
107-108	68	113	130	116	164	100	65	134	80	70	103	78	75	44	27	1367	50.5
105-106	81	121	148	117	136	82	36	89	57	61	31	66	42	31	11	1119	55.6
100-104	125	196	206	193	245	109	63	118	93	89	85	52	53	31	24	1682	63.2
95-99	250	293	323	226	249	140	81	125	83	100	96	45	39	31	17	2098	72.7
89-94	369	375	386	260	242	123	58	107	65	48	42	33	33	22	11	2174	32.5
81-88	348	373	324	223	176	79	50	64	42	36	22	24	15	5	3	1784	90.6
39-80	469	499	409	216	191	85	28	68	31	14	20	22	15	7	2	2076	100.0
TOTAL	1852	2220	2391	1797	2151	1201	751	1546	1133	1162	1242	1132	1297	1328	929	22132	
C-PCT	100.0	91.5	81.5	70.7	62.6	52.9	47.5	44.1	37.1	32.0	26.8	21.2	16.1	10.2	4.2		

Table 12

ACB-V and ACB-A Means and Standard Deviations for Grouped Data

ACB-V	ACB-A		ACB-A	ACB-V	
Interval	Mean	SD	Interval	Mean	SD
39-83	89.2	14.8	39-80	92.4	14.8
84-90	92.3	14.5	81-88	95.3	14.4
91-97	95.6	15.5	89-94	97.5	15.7
98-102	98.6	15.6	95-99	101.7	16.0
103-107	102.7	15.8	100-104	105.2	15.5
108-110	104.4	14.9	105-106	106.0	15.6
111-113	107.6	15.4	107-108	109.4	15.8
114-116	109.5	14.8	109-112	111.5	14.9
117-118	111.9	14.7	113-114	111.0	14.9
119-122	113.2	13.7	115-117	114.8	14.2
123-125	116.2	15.0	118-120	117.3	14.3
126-129	117.0	14.5	121-125	121.3	13.2
130-132	120.6	14.3	126-129	123.2	13.3
133-136	124.7	13.9	130-134	126.3	12.4
137-160	128.5	14.4	135-160	131.8	12.1

Table 13

Selected Comparisons of Extrapolated Correlations with Applicant Group Values for the PSAT-AIM Data with Selection Based on PSAT-V

Variable Correlated with Selection Variable	Correlation in Applicant Group	Correlations in Selected Group					Extrapolated Correlations				
		Percent in Selected Group					Percent in Selected Group				
		10	30	50	70	90	10	30	50	70	90
PSAT-Q	.75	.36	.49	.58	.65	.73	.66	.68	.71	.73	.75
English	.27	.11	.19	.24	.27	.28	.24	.30	.32	.33	.30
Music	.11	.02	.08	.11	.13	.13	.04	.14	.16	.16	.13
Soc. Sci.	.22	.11	.17	.20	.23	.23	.25	.28	.28	.28	.25
Math.	.15	.08	.11	.14	.15	.16	.17	.17	.19	.19	.17
Phy. Sci.	.14	.07	.14	.16	.17	.16	.16	.23	.22	.21	.17
Engr.	-.02	-.00	.00	.00	.00	-.01	-.01	-.01	.01	.00	.01
H. Econ.	-.12	-.13	-.12	-.14	-.13	-.11	-.27	-.20	-.19	-.16	-.12
Fine Arts	.06	-.04	.02	.02	.04	.06	-.09	.03	.03	.05	.06
Bio. Sci.	.05	.01	.04	.07	.07	.07	.01	.07	.09	.08	.07
Secretar.	-.24	-.15	-.23	-.24	-.27	-.25	-.31	-.36	-.33	-.32	-.26
For. Lang.	.22	.10	.15	.18	.21	.22	.22	.24	.24	.26	.24
Exect.	-.08	-.09	-.09	-.10	-.10	-.09	-.20	-.15	-.14	-.12	-.09

Table 14

Selected Comparisons of Extrapolated Correlations with

Applicant Group Values for the PSAT-AIM Data

with Selection Based on PSAT-V

Correlations between PSAT-Q and	Correlation in Applicant Group	Correlations in Selected Group					Extrapolated Correlations				
		Percent in Selected Group					Percent in Selected Group				
		10	30	50	70	90	10	30	50	70	90
English	.11	-.17	-.08	.00	.06	.11	-.01	-.19	.12	.14	.13
Music	.03	-.10	-.05	-.01	.02	.03	-.06	-.10	.05	.06	.05
Soc. Sci.	.13	-.04	.02	.07	.10	.13	.10	-.08	.15	.16	.15
Math.	.38	.49	.48	.47	.45	.41	.48	.49	.47	.45	.41
Phy. Sci.	.20	.21	.22	.24	.24	.22	.25	.18	.28	.27	.23
Engr.	.11	.19	.17	.17	.15	.13	.15	.19	.15	.14	.12
H. Econ.	-.19	-.21	-.20	-.21	-.21	-.19	-.31	-.16	-.24	-.23	-.20
Fine Arts	-.02	-.19	-.12	-.08	-.06	-.03	-.20	-.15	-.06	-.04	-.02
Bio. Sci.	.05	.03	.05	.07	.07	.07	.04	.04	.09	.09	.07
Secretar.	-.24	-.05	-.16	-.21	-.24	-.24	-.20	-.06	-.29	-.29	-.25
For. Lang.	.11	-.08	-.02	.03	.07	.11	.05	-.10	.11	.13	.12
Exect.	-.06	.05	-.01	-.05	-.06	-.06	-.07	.04	-.09	-.09	-.07

Table 15

Selected Comparisons of Extrapolated Correlations with Applicant Group Values for the Army Data with Selection Based on ACB-V

Variable Correlated with Selection Variable	Correlation in Applicant Group	Correlations in Selected Group						Extrapolated Correlations					
		Percent in Selected Group						Percent in Selected Group					
		10	20	40	60	80	90	10	20	40	60	80	90
ACB-A	.60	.16	.27	.38	.47	.55	.58	.42	.60	.64	.65	.65	.64
Pat. Anal.	.40	.09	.14	.23	.30	.37	.40	.26	.36	.43	.46	.46	.45
Mech. Apt.	.45	.10	.16	.26	.33	.40	.43	.27	.39	.48	.50	.49	.48
ACS	.36	.07	.10	.17	.24	.30	.33	.19	.25	.32	.37	.38	.38
Shop M	.31	.01	.06	.11	.18	.25	.28	.04	.15	.22	.29	.32	.33
Auto Inf.	.23	-.03	-.01	.03	.10	.16	.20	-.09	-.01	.06	.16	.21	.23
El Inf.	.40	.08	.12	.20	.28	.35	.37	.22	.30	.39	.43	.43	.42

Table 16

Selected Comparisons of Extrapolated Correlations with Applicant Group Values for the Army Data with Selection Based on ACB-V

Correlations between ACB-A and	Correlation in Applicant Group	Correlations in Selected Group						Extrapolated Correlations					
		Percent in Selected Group						Percent in Selected Group					
		10	20	40	60	80	90	10	20	40	60	80	90
Pat. Anal.	.51	.42	.45	.46	.49	.51	.51	.47	.54	.57	.57	.56	.54
Mech. Apt.	.48	.32	.37	.39	.42	.45	.47	.38	.49	.53	.53	.52	.51
ACS	.46	.38	.39	.38	.41	.44	.45	.41	.44	.46	.48	.49	.48
Shop M	.32	.17	.21	.22	.26	.29	.30	.17	.25	.29	.33	.34	.34
Auto Inf.	.26	.10	.14	.16	.19	.22	.24	.06	.10	.16	.23	.26	.26
EL Inf.	.40	.24	.28	.30	.33	.36	.38	.30	.38	.42	.44	.43	.42

Table 17

Comparison of Two New Correction Methods with

Standard Method Using PSAT Data

	Percent in Selected Group				
	10	30	50	70	90
PSAT-Q					
Applicant Group Correlation	.75	.75	.75	.75	.75
Standard Method	.66	.68	.71	.73	.75
New Method 1	.78	.78	.77	.78	.76
New Method 2	.64	.72	.75	.77	.76
English					
Applicant Group Correlation	.27	.27	.27	.27	.27
Standard Method	.24	.30	.32	.33	.30
New Method 1	.23	.29	.32	.33	.30
New Method 2	.21	.28	.31	.32	.29
Music					
Applicant Group Correlation	.11	.11	.11	.11	.11
Standard Method	.04	.14	.16	.16	.13
New Method 1	.04	.14	.16	.16	.13
New Method 2	.04	.14	.15	.16	.13
Soc. Sci.					
Applicant Group Correlation	.22	.22	.22	.22	.22
Standard Method	.25	.28	.28	.28	.25
New Method 1	.23	.28	.28	.28	.25
New Method 2	.21	.27	.27	.27	.24
Math					
Applicant Group Correlation	.15	.15	.15	.15	.15
Standard Method	.17	.17	.19	.19	.17
New Method 1	.17	.17	.19	.19	.17
New Method 2	.16	.17	.18	.19	.17
Phy. Sci.					
Applicant Group Correlation	.14	.14	.14	.14	.14
Standard Method	.16	.23	.22	.21	.17
New Method 1	.15	.22	.22	.21	.17
New Method 2	.15	.22	.22	.21	.17

Table 17 (continued)

	Percent in Selected Group				
	10	30	50	70	90
Engr.					
Applicant Group Correlation	-.02	-.02	-.02	-.02	-.02
Standard Method	-.01	-.01	.01	.00	-.01
New Method 1	-.01	-.01	.01	.00	-.01
New Method 2	.00	-.01	.01	.00	-.01
H. Econ.					
Applicant Group Correlation	-.12	-.12	-.12	-.12	-.12
Standard Method	-.27	-.20	-.19	-.16	-.12
New Method 1	-.28	-.20	-.19	-.16	-.12
New Method 2	-.27	-.20	-.19	-.16	-.12
Fine Arts					
Applicant Group Correlation	.06	.06	.06	.06	.06
Standard Method	-.09	.03	.03	.05	.06
New Method 1	-.09	.03	.03	.05	.06
New Method 2	-.09	.03	.03	.05	.06
Bio. Sci.					
Applicant Group Correlation	.05	.05	.05	.05	.05
Standard Method	.01	.07	.09	.08	.07
New Method 1	.01	.07	.09	.08	.07
New Method 2	.01	.07	.09	.08	.07
Secretar.					
Applicant Group Correlation	-.24	-.24	-.24	-.24	-.24
Standard Method	-.31	-.36	-.33	-.32	-.26
New Method 1	-.31	-.36	-.33	-.32	-.26
New Method 2	-.29	-.35	-.32	-.32	-.26
For. Lang.					
Applicant Group Correlation	.22	.22	.22	.22	.22
Standard Method	.22	.24	.24	.26	.24
New Method 1	.20	.23	.24	.25	.23
New Method 2	.18	.22	.23	.25	.23
Exect.					
Applicant Group Correlation	-.08	-.08	-.08	-.08	-.08
Standard Method	-.20	-.15	-.14	-.12	-.09
New Method 1	-.20	-.15	-.14	-.12	-.09
New Method 2	-.19	-.15	-.14	-.13	-.09

Table 18

Comparison of Two New Correction Methods with

Standard Method Using Army Data

	Percent in Selected Group				
	20	40	60	80	90
ACB-A					
Applicant Group Correlation	.60	.60	.60	.60	.60
Standard Method	.60	.64	.65	.65	.64
New Method 1	.58	.60	.58	.61	.63
New Method 2	.70	.73	.73	.71	.68
Pat. Anal.					
Applicant Group Correlation	.40	.40	.40	.40	.40
Standard Method	.36	.43	.46	.46	.45
New Method 1	.34	.41	.42	.44	.44
New Method 2	.37	.45	.48	.48	.46
Mech. Apt.					
Applicant Group Correlation	.45	.45	.45	.45	.45
Standard Method	.39	.48	.50	.49	.48
New Method 1	.46	.49	.48	.48	.48
New Method 2	.44	.55	.55	.52	.51
ACS					
Applicant Group Correlation	.36	.36	.36	.36	.36
Standard Method	.25	.32	.37	.38	.38
New Method 1	.28	.31	.34	.36	.37
New Method 2	.26	.33	.38	.39	.39
Shop M					
Applicant Group Correlation	.31	.31	.31	.31	.31
Standard Method	.15	.22	.29	.32	.33
New Method 1	.16	.24	.29	.32	.33
New Method 2	.16	.23	.31	.34	.34
Auto. Inf.					
Applicant Group Correlation	.23	.23	.23	.23	.23
Standard Method	-.01	.06	.16	.21	.23
New Method 1	-.02	.06	.16	.21	.23
New Method 2	-.02	.06	.16	.22	.23
El. Inf.					
Applicant Group Correlation	.40	.40	.40	.40	.40
Standard Method	.30	.39	.43	.43	.42
New Method 1	.41	.42	.41	.42	.42
New Method 2	.35	.44	.48	.47	.45

Table A

Univariate Distribution for ACB-V Scale

ACB-Verbal

Score	Freq.	Pct.	C-Pct.
50	72	0.3	0.3
51	0	0.0	0.3
52	0	0.0	0.3
53	0	0.0	0.3
54	0	0.0	0.3
55	25	0.1	0.4
56	3	0.0	0.5
57	2	0.0	0.5
58	0	0.0	0.5
59	0	0.0	0.5
60	31	0.1	0.6
61	0	0.0	0.6
62	19	0.1	0.7
63	0	0.0	0.7
64	20	0.1	0.8
65	28	0.1	0.9
66	1	0.0	0.9
67	31	0.1	1.0
68	77	0.3	1.4
69	30	0.1	1.5
70	37	0.2	1.7
71	119	0.5	2.2
72	53	0.2	2.5
73	50	0.2	2.7
74	128	0.6	3.3
75	50	0.2	3.5
76	59	0.3	3.8
77	119	0.5	4.3
78	65	0.3	4.6
79	73	0.3	4.9
80	362	1.6	6.6
81	3	0.0	6.6
82	394	1.8	8.3
83	1	0.0	8.4
84	409	1.8	10.2
85	110	0.5	10.7
86	259	1.2	11.9
87	134	0.6	12.5
88	609	2.7	15.2
89	5	0.0	15.2
90	694	3.1	18.4
91	155	0.7	19.1
92	657	3.0	22.0
93	5	0.0	22.1
94	507	2.3	24.3
95	137	0.6	25.0
96	74	0.3	25.3
97	859	3.9	29.2
98	257	1.2	30.3
99	200	0.9	31.2
100	1149	5.2	36.4

Table A (Continued)

ACB-Verbal

Score	Freq.	Pct.	C-Pct.
101	2	0.0	36.4
102	190	0.9	37.3
103	988	4.5	41.7
104	5	0.0	41.8
105	256	1.2	42.9
106	679	3.1	46.0
107	227	1.0	47.0
108	5	0.1	47.0
109	245	1.1	48.1
110	952	4.3	52.4
111	308	1.4	53.8
112	142	0.6	54.4
113	302	1.4	55.8
114	1011	4.6	60.4
115	151	0.7	61.0
116	387	1.7	62.8
117	301	1.4	64.1
118	833	3.8	67.9
119	449	2.0	69.9
120	16	0.1	70.0
121	554	2.5	72.5
122	144	0.6	73.2
123	512	2.3	75.5
124	3	0.0	75.5
125	733	3.3	78.8
126	716	3.2	82.0
127	8	0.0	82.0
128	407	1.8	83.9
129	2	0.0	83.9
130	1298	5.9	89.7
131	3	0.0	89.8
132	3	0.0	89.8
133	710	3.2	93.0
134	1	0.0	93.0
135	1	0.0	93.0
136	621	2.8	95.8
137	3	0.0	95.8
138	6	0.0	95.8
139	2	0.0	95.8
140	275	1.2	97.1
141	1	0.0	97.1
142	0	0.0	97.1
143	4	0.0	97.1
144	0	0.0	97.1
145	364	1.6	98.7

Table A (Continued)

ACB-Verbal

Score	Freq.	Pct.	C-Pct.
146	1	0.0	98.7
147	1	0.0	98.8
148	1	0.0	98.8
149	0	0.0	98.8
150	1	0.0	98.8
151	0	0.0	98.8
152	276	1.2	100.0

Mean	108.89
SD	18.61
Coef Skewness	-.14
Coef Kurtosis	-.36

Table B

Univariate Distribution for ACB-A Scale

ACB-Arithmetic

Score	Freq.	Pct.	C-Pct.
50	121	0.5	0.5
51	0	0.0	0.5
52	0	0.0	0.5
53	0	0.0	0.5
54	0	0.0	0.5
55	27	0.1	0.7
56	0	0.0	0.7
57	0	0.0	0.7
58	32	0.1	0.8
59	0	0.0	0.8
60	54	0.2	1.1
61	0	0.0	1.1
62	1	0.0	1.1
63	1	0.0	1.1
64	113	0.5	1.6
65	1	0.0	1.6
66	4	0.0	1.6
67	0	0.0	1.6
68	209	0.9	2.5
69	0	0.0	2.5
70	4	0.0	2.6
71	95	0.4	3.0
72	201	0.9	3.9
73	0	0.0	3.9
74	130	0.6	4.5
75	218	1.0	5.5
76	137	0.6	6.1
77	9	0.0	6.1
78	171	0.8	6.9
79	147	0.7	7.6
80	401	1.8	9.4
81	186	0.8	10.2
82	455	2.1	12.3
83	2	0.0	12.3
84	580	2.6	14.9
85	3	0.0	14.9
86	2.3	1.0	15.9
87	266	1.2	17.1
88	79	0.4	17.4
89	238	1.1	18.5
90	645	2.9	21.4

Table B (Continued)

ACB-Arithmetic

Score	Freq.	Pct.	C-Pct.
91	303	1.4	22.8
92	72	0.3	23.1
93	882	4.0	27.1
94	36	9.2	27.2
95	245	1.1	28.3
96	675	3.0	31.4
97	366	1.7	33.0
98	141	0.6	33.7
99	675	3.0	36.7
100	158	0.7	37.4
101	350	1.6	39.0
102	817	3.7	42.7
103	356	1.6	44.3
104	4	0.0	44.3
105	1110	5.0	49.3
106	10	0.0	49.4
107	573	2.6	51.9
108	797	3.6	55.5
109	9	0.0	55.6
110	385	1.7	57.3
111	463	2.1	59.4
112	341	1.5	60.9
113	9	0.0	61.0
114	1012	4.6	65.5
115	174	0.8	66.3
116	430	1.9	68.3
117	664	3.0	71.3
118	459	2.1	73.3
119	5	0.0	73.4
120	976	4.4	77.8
121	7	0.0	77.8
122	415	1.9	79.7
123	150	0.7	80.3
124	9	0.0	80.4
125	697	3.1	83.5
126	487	2.2	85.7
127	384	1.7	87.4
128	10	0.0	87.5
129	374	1.7	89.2
130	575	2.6	91.8
131	393	1.8	93.5

Table B (Continued)

ACB-Arithmetic

Score	Freq.	Pct	C-Pct.
132	0	0.0	93.5
133	3	0.0	93.6
134	312	1.4	95.0
135	3	0.0	95.0
136	294	1.1	96.1
137	2	0.0	96.1
138	188	0.8	97.0
139	1	0.0	97.0
140	3	0.0	97.0
141	207	0.9	97.9
142	0	0.0	97.9
143	153	0.7	98.6
144	1	0.0	98.6
145	4	0.0	98.6
146	119	0.5	99.2
147	1	0.0	99.2
148	1	0.0	99.2
149	70	0.3	99.5
150	0	0.0	99.5
151	0	0.0	99.5
152	55	0.2	99.7
153	0	0.0	99.7
154	0	0.0	99.7
155	0	0.0	99.7
156	0	0.0	99.7
157	0	0.0	99.7
158	0	0.0	99.7
159	0	0.0	99.7
160	59	0.3	100.0

Mean	106.14
SD	18.78
Coef. Skewness	-.15
Coef. Kurtosis	-.21

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1 ORIGINATING ACTIVITY (Corporate author) Educational Testing Service Princeton, New Jersey 08540	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3 REPORT TITLE

AN INVESTIGATION OF THE ACCURACY OF THE PEARSON SELECTION FORMULAS

4 DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technica Report

5. AUTHOR(S) (First name, middle initial, last name)

Melvin R. Novick and Dorothy T. Thayer

6 REPORT DATE September 1969	7a. TOTAL NO OF PAGES 46	7b. NO OF REFS 1
8a. CONTRACT OR GRANT NO. 000-14-69C-0119 b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) RM-69-22 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	

10 DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11 SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Office of Naval Research Navy Department Washington, D. C. 20360

13. ABSTRACT

The Pearson formulas for correcting correlation coefficients for restriction of range are based on crucial assumptions of linearity of regression and homoscedasticity of the error distributions. Some small previous studies have suggested that these formulas are reasonably accurate providing extreme selection is not involved. These studies tend to suggest that the formulas typically provide undercorrections in most instances. The present study involved two very large data sets and attempted to verify the accuracy or inaccuracy of these formulas and the assumptions on which they are based for both moderate and extreme selection. Generally, it was found that the linearity assumption was reasonably well satisfied except in the extreme tails of the distribution while the homoscedasticity assumption was not. In neither set of data did the correction formulas work as well as previous research had led the authors to expect they would. Undoubtedly this was due to the invalidity of the homoscedasticity assumption. Some methods for taking into account heteroscedasticity of errors were studied and some very minor improvements were found. However, no method seems to have any general validity.

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6801

Security Classification

3ND PPSO 13152

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
range restriction						
explicit selection						
incidental selection						
correlation						
regression						

DD FORM 1 NOV 65 1473 (BACK)
(PAGE 2)

Security Classification